10

15

20



1. A method of determining a title from a document image, comprising:
dividing the document image into minimal circumscribing rectangles which
contain a character image,

recognizing characters in said minimal circumscribing rectangles; and determining a title of the document image based upon a likelihood of each of said minimal circumscribing rectangles containing a title, said likelihood being determined based upon information obtained during said character recognition.

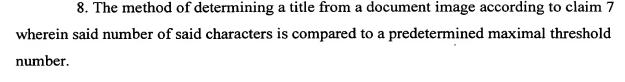
- 2. The method of determining a title from a document image according to claim 1 wherein said likelihood is expressed in a sum of points based said information.
- 3. The method of determining a title from a document image according to claim 2 wherein said information includes characteristics on font.
- 4. The method of determining a title from a document image according to claim 2 wherein said font characteristics is determined on a frequency of a particular font type.
- 5. The method of determining a title from a document image according to claim 2 wherein said character recognition further includes an act of matching said characters with a set of predetermined words, said predetermined words indicating said title.
- 25 6. The method of determining a title from a document image according to claim 5 wherein said information includes a result of said matching with said predetermined words.
 - 7. The method of determining a title from a document image according to claim 2 wherein said information includes a number of said characters.

15

20

25

5



wherein said information includes an assurance level of said character recognition.

10. The method of determining a title from a document image according to claim
9 wherein said assurance level is compared to a predetermined minimal threshold value.

11. The method of determining a title from a document image according to claim 2 wherein said information includes layout characteristics.

- 12. The method of determining a title from a document image according to claim 11 wherein said information includes centering, underlining and size.
- 13. The method of determining a title from a document image according to claim 2 wherein said information indicates whether or not said characters end in a noun form.
- 14. The method of determining a title from a document image according to claim 2 wherein said information indicates whether or not said characters end in a set of predetermined suffixes.
- 15. The method of determining a title from a document image according to claim 2 wherein said information includes a ratio between a length and a height of each of said circumscribing rectangles.

- 16. The method of determining a title from a document image according to claim 2 wherein said information includes a ratio between a summed width of said characters and a corresponding one of said circumscribing rectangles.
- 17. The method of determining a title from a document image according to claim 1 wherein said likelihood is adjusted according to a type of said image documents.
- 18. The method of determining a title from a document image according to claim 1 wherein said title is combined with a keyword.

19. A system for determining a title from a document image, comprising: a character row area determination unit for dividing the document image into minimal circumscribing rectangles which contain a character image;

a character recognition unit connected to said character row area determination unit for recognizing characters in said minimal circumscribing rectangles; and

a title evaluation point determination unit connected to said character recognition unit for determining a title of the document image based upon a likelihood of each of said minimal circumscribing rectangles containing a title, said likelihood being determined based upon information obtained during said character recognition.

20. The system for determining a little from a document image according to claim 19 wherein said title evaluation point determination unit determines said likelihood in terms of a sum of points based said information.

21. The system for determining a title from a document image according to claim 20 wherein said title evaluation point determination unit further comprises a font determination unit for generating information on font characteristics.

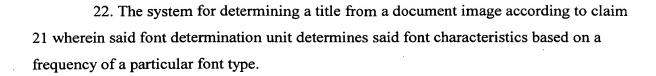
20

15

5

10

25



- 23. The system for determining a title from a document image according to claim 20 wherein said title evaluation point determination unit further comprises a natural language analysis unit for matching said characters with a set of predetermined words, said predetermined words indicating said title.
- 24. The system for determining a title from a document image according to claim 23 wherein said natural language analysis unit generates a result of matching of said characters with said predetermined words.

25. The system for determining a title from a document image according to claim 20 wherein said character recognition unit generates said information on a number of said characters.

26. The system for determining a title from a document image according to claim 25 wherein said title evaluation point determination unit compares said number of said characters to a predetermined maximal threshold number.

27. The system for determining a title from a document image according to claim 20 wherein said character recognition unit generates said information on an assurance level of said character recognition.

25

20

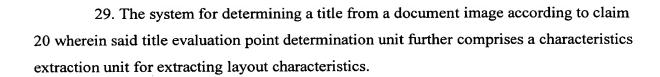
5

28. The system for determining a title from a document image according to claim 27 wherein said title evaluation point determination unit compares said assurance level to a predetermined minimal threshold value.

15

20

25



- 5 30. The system for determining a title from a document image according to claim 29 wherein said extraction unit extracts said layout characteristics on centering, underlining and size.
- 31. The system for determining a title from a document image according to claim
 23 wherein said natural language analysis unit generates said information indicating
 whether or not said characters end in a noun form.
 - 32. The system for determining a title from a document image according to claim 23 wherein said natural language analysis unit generates said information indicating whether or not said characters end in a set of predetermined suffixes.
 - 33. The system for determining a title from a document image according to claim 2 wherein said character row area determination unit generates said information on a ratio between a length and a height of each of said circumscribing rectangles.
 - 34. The system for determining a title from a document image according to claim 2 wherein said character row area determination unit generates said information on a ratio between a summed width of said characters and a corresponding one of said circumscribing rectangles.
 - 35. The system for determining a title from a document image according to claim 19 wherein said likelihood is adjusted according to a type of said image documents.
- 36. The system for determining a title from a document image according to claim19 wherein said title is combined with a keyword.